TELEPATHY LABS

Problem

A customer service department has been receiving questions to which answers have been provided up to now manually. Over time, the team has realized that the amount of questions to be answered has continuously grown to a number that cannot be handled manually anymore with the given resources; but it has also realized that the answering to the questions could be automated.

The customer service has a database with a vast amount of questions. Each question in textual form (T_n) is mapped in the database to a unique answer ID (A_n).

To implement the automatic answering system, the engineering team has proposed to implement a NLU-based approach where intents and entities are derived from the questions to implement machine understandable queries (Q_n). To implement that, the engineering team proposes to build a new mapping table: (Q_n) → (A_n). A NLU system should be trained, and once in production, the user question, arriving in a textual form (T_u) should be transformed by this system into the query (Q_u) that is in turn used to provide the answer (A_u) using the mapping table.

A subset of the questions {T_n} is provided to you (n=1..1500). The majority of nouns in the text have been replaced by a code NNy, or NNSz, where y, z are integers.

You should invest a maximum of 8 hours in answering to the following questions and implementing the task. After this time, please sends us the outcome of your efforts via email.

Questions & Tasks

Q1. How do you find the engineering proposal? Would your solution be similar? Or would you follow a different approach? Please explain your answer and if you would follow a different approach, please describe it.

Q2. Do you see any problem with the current available data, i.e. the table (T_n) → (A_n) to achieve the desired automated system?

T1. With the data provided, find* an adequate (Q_n) set. i.e. find the set of intents, entities and entities values that are needed so that a trained NLU system can reliably transform the variability present in the "natural language" users' input (T_u) into a normalized query (Q_u) that can lead to the answer (A_u).

* Prototype an algorithm using python 3.x and any additional open source library that you may require.

Q3. Describe a possible evaluation method for the task T1.

T2. (Optional) Use the evaluation method you proposed in Q3 to evaluate your solution to T1.